STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN 2053-230X

Structure of uridine diphosphate *N*-acetylglucosamine pyrophosphorylase from *Entamoeba histolytica*

Thomas E. Edwards,^{a,b*} Anna S. Gardberg,^{a,b‡} Isabelle Q. H. Phan,^{a,c} Yang
Zhang,^{a,c} Bart L. Staker,^{a,c} Peter J. Myler^{a,c,d} and Donald D. Lorimer^{a,b}

Received 1 December 2014

Accepted 27 January 2015

Edited by W. N. Hunter, University of Dundee,
Scotland‡ Present address: EMD Serono Research and
Development Institute Inc., Billerica,
MA 01821, USA.**Keywords:** SSGCID; *Entamoeba histolytica*;
amoebic dysentery; parasitic protozoan;
allosteric regulation; cell-wall biogenesis;
GlcNAc.**PDB reference:** uridine diphosphate *N*-acetyl-
glucosamine pyrophosphorylase, 3oc9**Supporting information:** this article has
supporting information at journals.iucr.org/f^aSeattle Structural Genomics Center for Infectious Disease, USA, ^bBeryllium, Bainbridge Island, WA 98110, USA, ^cSeattle
Biomedical Research Institute, Seattle, WA 98109, USA, and ^dDepartments of Global Health and Medical Education and
Biomedical Informatics, University of Washington, Seattle, WA 98195, USA. *Correspondence e-mail:
tedwards@be4.com

Uridine diphosphate *N*-acetylglucosamine pyrophosphorylase (UAP) catalyzes the final step in the synthesis of UDP-GlcNAc, which is involved in cell-wall biogenesis in plants and fungi and in protein glycosylation. Small-molecule inhibitors have been developed against UAP from *Trypanosoma brucei* that target an allosteric pocket to provide selectivity over the human enzyme. A 1.8 Å resolution crystal structure was determined of UAP from *Entamoeba histolytica*, an anaerobic parasitic protozoan that causes amoebic dysentery. Although *E. histolytica* UAP exhibits the same three-domain global architecture as other UAPs, it appears to lack three α -helices at the N-terminus and contains two amino acids in the allosteric pocket that make it appear more like the enzyme from the human host than that from the other parasite *T. brucei*. Thus, allosteric inhibitors of *T. brucei* UAP are unlikely to target *Entamoeba* UAPs.

1. Introduction

Uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) is an essential metabolite for the generation of peptidoglycans used in cell-wall synthesis in fungi and plants, and is used in protein glycosylation in other eukaryotes (Peneff *et al.*, 2001). UDP-GlcNAc is the final product of the hexosamine biosynthetic pathway, which starts with the conversion of glucose to glucose 6-phosphate by hexokinase (Aleshin, Zeng, Bartunik *et al.*, 1998; Aleshin, Zeng, Bourenkov *et al.*, 1998), includes the addition of the amine *via* the conversion of fructose 6-phosphate to glucosamine 6-phosphate by GFAT (Nakaishi *et al.*, 2009), and ends in the conversion of UTP and *N*-acetylglucosamine 1-phosphate (GlcNAc-1-P) to UDP-GlcNAc and inorganic pyrophosphate by uridine diphosphate *N*-acetylglucosamine pyrophosphorylase (UAP; Peneff *et al.*, 2001). Mechanistic understanding and the associated conformational changes during the conversion of GlcNAc-1-P to UDP-GlcNAc have been derived for UAP from the fungus *Candida albicans* (Maruyama *et al.*, 2007) in comparison to the human (Peneff *et al.*, 2001) and mouse (PDB entry 1vm8; Joint Center for Structural Genomics, unpublished work) enzymes. In prokaryotes, the same enzymatic reaction is carried out by the related enzyme GlmU, a bifunctional, multi-domain enzyme which both acetylates glucosamine 1-phosphate to GlcNAc-1-P and converts GlcNAc-1-P to UDP-GlcNAc.

© 2015 International Union of Crystallography

Table 1

Data-collection and processing statistics for *EhUAP*.

Values in parentheses are for the outer shell.

Wavelength (Å)	1.000
Crystal-to-detector distance (mm)	250
Rotation range per image (°)	1.0
Total rotation range (°)	240
Exposure time per image (s)	3
Space group	$P2_12_12_1$
a, b, c (Å)	76.52, 77.54, 86.87
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.3
Resolution range (Å)	50–1.80 (1.84–1.80)
Total No. of reflections	146263 (10489)
No. of unique reflections	45589 (3573)
Completeness (%)	100 (100)
Multiplicity	9.7 (9.1)
$\langle I/\sigma(I) \rangle$	30.5 (4.4)
R_{merge}	0.046 (0.529)
R_{meas}	0.049 (0.561)
Overall B factor from Wilson plot (Å ²)	31.9

Enzymes involved in cell-wall synthesis have long been the target of antimicrobials because of the importance of cell-wall homeostasis and the fact that antimicrobial agents do not necessarily need to cross the cell wall to be effective therapeutics. Recently, UAP from *Aspergillus fumigatus* was shown to be a valid antifungal target (Fang *et al.*, 2013), and an allosteric inhibitor of *Trypanosoma brucei* UAP has been reported (Urbaniak *et al.*, 2013). Similarly, inhibitors have been described for the bacterial bifunctional GlmU enzymes analogous to UAP that target the pyrophosphorylase domain (Green *et al.*, 2012; Larsen *et al.*, 2012; Mochalkin *et al.*, 2007, 2008; Stokes *et al.*, 2012). Therefore, we sought to structurally characterize the UAP from *Entamoeba histolytica* (*EhUAP*), the parasitic protozoan responsible for amoebic dysentery, in order to understand whether or not *EhUAP* represents a suitable drug target and whether or not the reported *T. brucei* UAP inhibitor (Urbaniak *et al.*, 2013) may also target *E. histolytica*. *E. histolytica* contains two copies of UAP (EnhiA.01126.a, UniProt ID C4MA87, and EnhiA.01126.b, UniProt ID C4M036) that exhibit 96% sequence identity to one another. Both are listed as potential drug targets in EuPathDB (AmoebaDB:EHI_039830 and AmoebaDB:EHI_021200, respectively).

2. Materials and methods

2.1. Protein expression and purification

The 401-residue *E. histolytica* UAP gene (UniProt accession code C4M036) was amplified from genomic DNA and cloned into an expression vector (pAVA0421) encoding an N-terminal hexahistidine affinity tag followed by the human rhinovirus 3C protease cleavage sequence using ligation-independent cloning (Aslanidis & de Jong, 1990). The sequence of the entire tag is MAHHHHHHMGTLEAQT-QGPS, which is followed by the 401-residue *E. histolytica* UAP. Although *EhUAP* was cloned from genomic DNA, two amino acids were different in the clone in comparison with the sequence reported in the database: L37P and M306I. After

Table 2

Structure solution and refinement of *EhUAP*.

Values in parentheses are for the outer shell.

Resolution range (Å)	50–1.80 (1.85–1.80)
Completeness (%)	99.7
σ Cutoff	$F > 0.000\sigma(F)$
No. of reflections, working set	46018 (3358)
No. of reflections, test set	2450 (202)
Final R_{cryst}	0.191 (0.283)
Final R_{free}	0.227 (0.306)
No. of non-H atoms	
Protein	3199
Water	380
Total	3588
R.m.s. deviations	
Bonds (Å)	0.018
Angles (°)	1.300
Average B factors (Å ²)	
Protein	31.3
Water	37.4
Ramachandran plot	
Most favoured (%)	97.4
Allowed (%)	100
PDB code	3oc9

transformation into *Escherichia coli* BL21 (DE3) cells, starter cultures of LB broth were grown for ~18 h at 37°C. Protein was expressed in a LEX bioreactor in the presence of antibiotics in 2 l sterilized ZYP-5052 auto-induction medium (Studier, 2005) inoculated with the overnight starter culture. After 24 h at 25°C, the temperature was reduced to 15°C for a further 60 h. The sample was centrifuged at 4000g for 20 min at 4°C. The cell paste was flash-frozen in liquid nitrogen and stored at –80°C. The frozen cells were resuspended in 25 mM HEPES pH 7.0, 500 mM NaCl, 5% (v/v) glycerol, 0.5% (w/v) CHAPS, 30 mM imidazole, 10 mM MgCl₂, 1 mM TCEP, 250 µg ml^{–1} AEBSEF, 0.025% (w/v) azide at 4°C. Lysis was achieved by sonication followed by incubation with Benzonase (20 µl at 25 units µl^{–1}). Insoluble proteins and other cellular components were removed by centrifugation at 4000g for 75 min at 4°C. The soluble fraction was loaded onto an Ni-NTA HisTrap FF 5 ml column (GE Healthcare). The column was washed with 20 column volumes of wash buffer [25 mM HEPES pH 7.0, 500 mM NaCl, 5% (v/v) glycerol, 30 mM imidazole, 1 mM TCEP, 0.025% (w/v) azide]. The bound protein was eluted with seven column volumes of elution buffer [25 mM HEPES pH 7.0, 500 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, 250 mM imidazole, 0.025% (w/v) azide]. Cleavage of the N-terminal His tag was accomplished by dialysis with His-MBP-3C protease at 4°C overnight in 25 mM HEPES pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, 0.025% (w/v) azide. The cleaved protein was recovered in the flowthrough and wash fractions of a subtractive Ni²⁺-affinity chromatography step that removed His-MBP-3C protease, uncleaved protein and the cleaved His tag. The collected cleaved protein (sequence GPGS followed by 401-residue *EhUAP*) was loaded onto a Hiload 26/60 Superdex 75 prep-grade column (GE Healthcare) equilibrated in 25 mM HEPES pH 7.0, 500 mM NaCl, 5% (v/v) glycerol, 2 mM dithiothreitol, 0.025% (w/v) azide. The purified protein was concentrated to 96.3 mg ml^{–1} and stored at –80°C.

Figure 1

Clustal Omega multiple sequence alignment of UAP enzymes from *Homo sapiens* (HsAGX1 and HsAGX2), the yeast *C. albicans* (CaUAP1), the filamentous fungus *A. fumigatus* (AfUAP) and the parasites *E. histolytica* (EhUAP) and *T. brucei* (TbUAP). *E. histolytica* contains two copies of the enzyme (UniProt IDs C4MA87 and C4M036). The structure presented here was solved from C4M036 (EhUAP_EHI_021200) and contains two amino-acid changes L37P and M306I. Residues are colored according to conservation (cyan is lowest, green is moderate and red is complete conservation). Active-site residues equivalent to those within 5 Å of UDP-GlcNAc as reported in the human AGX1 structure (PDB entry 1jv1) are indicated in yellow at the bottom of the alignment.

2.2. Crystallization

EhUAP crystallized at 96.3 mg ml^{-1} using JCSG+ screen condition H9, which consists of 0.2 M lithium sulfate, 0.1 M bis-tris pH 5.5, $25\%(\text{w/v})$ PEG 3350. Equal volumes of protein solution and precipitant ($0.4 \mu\text{l}$) were set up at 16°C against reservoir ($80 \mu\text{l}$) in sitting-drop vapor-diffusion format. Crystals appeared within a week and exhibited a brick-shaped morphology, often with irregular edges.

2.3. Data collection and processing

Crystals were cryoprotected with reservoir supplemented with $15\%(\text{v/v})$ ethylene glycol and flash-cooled in liquid nitrogen. A data set was collected at 100 K from an apo crystal under a stream of liquid nitrogen on beamline 5.0.3 at the Advanced Light Source (ALS; Table 1). Data were reduced with *XDS/XSCALE* (Kabsch, 2010).

2.4. Structure solution and refinement

The apo structure (Table 2) was solved by molecular replacement using residues 68–407 of human UAP isoform 1 (PDB entry 1jv1; Peneff *et al.*, 2001) as a search model in *Phaser* (McCoy *et al.*, 2007) from the *CCP4* suite (Winn *et al.*, 2011). The structure was rebuilt using *ARP/wARP* (Langer *et al.*, 2008). The final models were obtained after numerous rounds of refinement in *REFMAC* (Murshudov *et al.*, 2011) and manual rebuilding in *Coot* (Emsley & Cowtan, 2004). Structures were assessed for correctness and validated using *MolProbity* (Chen *et al.*, 2010).

3. Results and discussion

3.1. Apo *EhUAP* crystal structure

A $1.8 \text{ \AA}$ resolution crystal structure was determined for full-length *EhUAP* by molecular replacement using the human isoform 1 UAP (Peneff *et al.*, 2001) as a search model. Mole-

Figure 2

(a) Crystal structure of *E. histolytica* UAP showing the N-terminal domain in orange, the central pyrophosphorylase domain in green, the nucleotide-binding loop in magenta and the C-terminal domain in blue. (b) Overlay of *E. histolytica* UAP with the UDP-GlcNAc-bound and sulfate-bound structure of UAP from the yeast *C. albicans* (shown in gray). (c) Comparison of the active-site residues which are not conserved in *EhUAP*: Pro37 (normally a Met) and Ser43 (normally a Thr). Arrows indicate the movement of an active-site helix and loop which contract in the ligand-bound state, as observed in the *C. albicans* UAP structure. (d) Overlay of *E. histolytica* UAP shown as a space-filling model with the UDP-GlcNAc-bound structure of *C. albicans* UAP, highlighting the N-terminal α -helical extension present in other UAP enzymes.

cular replacement was only successful with a heavily truncated search probe consisting almost entirely of the core pyrophosphorylase domain. *Eh*UAP displays 33–38% sequence identity over 320–395 amino acids (Fig. 1) and r.m.s.d. values of 1.73–1.94 Å compared with the human isoform 1 and 2 UAP enzymes and those from mouse, the yeast *C. albicans*, the fungus *A. fumigatus* and the parasite *T. brucei*. Like other eukaryotic UAP enzymes, *Eh*UAP generally consists of three domains: an N-terminal domain, a central pyrophosphorylase domain and a C-terminal domain (Fig. 2*a*). Like UAP enzymes from other non-mammalian organisms, *Eh*UAP lacks the insertion loop between β 18 and β 19 of the C-terminal domain, which is involved in dimerization of the human enzyme, as well as the short C-terminal extension. Consequently, *Eh*UAP displays a monomeric quaternary structure in its crystalline state, consistent with other non-mammalian UAP enzyme architectures.

3.2. Comparison with apo and substrate-bound states

Co-crystallization of *Eh*UAP in the presence of UDP-GlcNAc, UDP-Glu, UDP-Gal, UDP or UMPPNP yielded crystals which diffracted to high resolution (1.9–2.1 Å) but failed to yield ligand-bound assemblies. Therefore, our understanding of substrate and product binding to *Eh*UAP is currently limited to comparison of overlays with ligand-bound structures of UAP from other species. The nucleotide-binding motif (residues 37–51) remains highly conserved in *Eh*UAP in comparison with other UAPs, and many of these structural features overlay with the equivalent features in the apo structures from other organisms. In *C. albicans* UAP this loop moves upon ligand binding (Fig. 2*b*). There are a couple of noteworthy changes in the nucleotide-binding motif sequence of *Eh*UAP (Fig. 2*c*). Firstly, the equivalent residue to Ser43 in *Eh*UAP is a threonine in almost all other organisms. This amino acid may be involved in the release of the inorganic pyrophosphate byproduct, as demonstrated by a product-state assembly of a yeast UAP in complex with UDP-GlcNAc and sulfate (Maruyama *et al.*, 2007), although it is not clear whether the change from threonine to serine would be detrimental here. The second amino-acid change, which is the replacement of the large hydrophobic amino-acid leucine or methionine with Pro37, may have a more significant effect on catalysis. In other structures the equivalent methionine residue is located under the sugar and uridine sugar edge, and it seems unlikely that a proline residue could make similar hydrophobic packing interactions. Furthermore, a proline residue at this position may alter the ability of the nucleotide-binding loop to adopt different conformations during substrate binding and catalysis. Although our construct was cloned from genomic DNA, residue 37 of *Eh*UAP is reported in UniProt to be a leucine. Thus, the proline at position 37 may interfere with ligand binding and/or catalysis, although it appears to have little effect on the global *Eh*UAP structure in comparison with other UAP enzymes (Fig. 2*b*). Other than these two point mutants and the expected opening of the nucleotide-binding loop in the absence of ligand (Fig. 2*c*), the

*Eh*UAP binding pocket largely appears consistent with other UAPs and is likely to follow the common UAP mechanism described for *C. albicans* UAP (Maruyama *et al.*, 2007).

3.3. Potential for allosteric inhibition

The active sites of UAP enzymes are highly conserved between human and parasites such as *T. brucei* or *E. histolytica*, and thus active-site inhibitors for *Entamoeba* UAPs would be likely to suffer from a lack of specificity over the human enzyme. Allosteric inhibitors of the parasitic proto-

Figure 3
(*a*) Comparison of the crystal structures of *T. brucei* UAP (gray ribbons) bound to an allosteric inhibitor (gray sticks) with *E. histolytica* UAP colored as in Fig. 1. The inhibitor binds to the back side of the interface of the N-terminal domain and pyrophosphorylase domain relative to UDP-GlcNAc. (*b*) Comparison of the allosteric inhibitor-binding site from *T. brucei* UAP colored gray, *E. histolytica* UAP colored as in Fig. 1 and human UAP shown as yellow ribbons with a yellow carbon backbone. Selectivity over the human enzyme is achieved in part by two residues. Ile306 in *E. histolytica* UAP is a lysine in the human enzyme, but an alanine in that from *T. brucei*. Asn151 is an aspartic acid in the human enzyme, but a glycine in that from *T. brucei*. The presence of larger amino-acid side chains at these positions indicates that allosteric inhibitors developed against the parasite *T. brucei* are unlikely to be effective against the parasite *E. histolytica*.

zoan *T. brucei* UAP have been developed that are selective over the human UAP enzyme (Urbaniak *et al.*, 2013). Unfortunately, it appears unlikely that these inhibitors will also be effective against the parasite protozoan *E. histolytica*. These allosteric inhibitors bind to the opposite side of the enzyme relative to the catalytic site at the interface between the N-terminal domain and the pyrophosphorylase domain. *EhUAP* lacks the three α -helices of the N-terminal extension (Fig. 2*d*) which comprise part of the binding site (Fig. 3*a*). Furthermore, the two key residues which provide specificity for *T. brucei* UAP over the human enzyme are more human-like in *EhUAP* (Fig. 3*b*). Asn151 is an aspartic acid in human UAP but is a glycine in *T. brucei* UAP. Likewise, residue 306 is an isoleucine in our *EhUAP* structure but is a lysine in human UAP and an alanine in *T. brucei* UAP. This residue is reported to be a methionine in both *EhUAP* genes and although our protein was cloned from genomic DNA, it contains an isoleucine at this position which may be a natural variant. In both *E. dispar* and *E. invadens* UAP these residues are Asn151 and Met306. Interestingly, the closest homolog in primary sequence to *Entamoeba* UAPs is the UAP from the yellow fever mosquito *Aedes aegypti*, which carries an aspartic acid and a lysine, respectively, at these positions, exactly the same as in humans. In conclusion, the development of *Entamoeba* UAP inhibitors would need to take a different approach from the allosteric inhibitor scaffold described for *T. brucei*.

Acknowledgements

This project has been funded in whole with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under Federal Contract Nos. HHSN272201200025C and HHSN272200700057C. The expression plasmid can be obtained from either the contact author or through <http://www.ssgcid.org>. Raw X-ray diffraction data can be obtained through www.csgid.org.

References

- Aleshin, A. E., Zeng, C., Bartunik, H. D., Fromm, H. J. & Honzatko, R. B. (1998). *J. Mol. Biol.* **282**, 345–357.
- Aleshin, A. E., Zeng, C., Bourenkov, G. P., Bartunik, H. D., Fromm, H. J. & Honzatko, R. B. (1998). *Structure*, **6**, 39–50.
- Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Fang, W., Du, T., Raimi, O. G., Hurtado-Guerrero, R., Urbaniak, M. D., Ibrahim, A. F. M., Ferguson, M. A., Jin, C. & van Aalten, D. M. F. (2013). *Mol. Microbiol.* **89**, 479–493.
- Green, O. M., McKenzie, A. R., Shapiro, A. B., Otterbein, L., Ni, H., Patten, A., Stokes, S., Albert, R., Kawatkar, S. & Breed, J. (2012). *Bioorg. Med. Chem. Lett.* **22**, 1510–1519.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Larsen, N. A., Nash, T. J., Morningstar, M., Shapiro, A. B., Joubran, C., Blackett, C. J., Patten, A. D., Boriack-Sjodin, P. A. & Doig, P. (2012). *Biochem. J.* **446**, 405–413.
- Maruyama, D., Nishitani, Y., Nonaka, T., Kita, A., Fukami, T. A., Mio, T., Yamada-Okabe, H., Yamada-Okabe, T. & Miki, K. (2007). *J. Biol. Chem.* **282**, 17221–17230.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mochalkin, I., Lightle, S., Narasimhan, L., Bornemeier, D., Melnick, M., VanderRoest, S. & McDowell, L. (2008). *Protein Sci.* **17**, 577–582.
- Mochalkin, I., Lightle, S., Zhu, Y., Ohren, J. F., Spessard, C., Chirgadze, N. Y., Banotai, C., Melnick, M. & McDowell, L. (2007). *Protein Sci.* **16**, 2657–2666.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nakaishi, Y., Bando, M., Shimizu, H., Watanabe, K., Goto, F., Tsuge, H., Kondo, K. & Komatsu, M. (2009). *FEBS Lett.* **583**, 163–167.
- Penef, C., Ferrari, P., Charrier, V., Taburet, Y., Monnier, C., Zamboni, V., Winter, J., Harnois, M., Fassy, F. & Bourne, Y. (2001). *EMBO J.* **20**, 6191–6202.
- Stokes, S. S., Albert, R., Buurman, E. T., Andrews, B., Shapiro, A. B., Green, O. M., McKenzie, A. R. & Otterbein, L. R. (2012). *Bioorg. Med. Chem. Lett.* **22**, 7019–7023.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Urbaniak, M. D., Collie, I. T., Fang, W., Aristotelous, T., Eskilsson, S., Raimi, O. G., Harrison, J., Navratilova, I. H., Frearson, J. A., van Aalten, D. M. & Ferguson, M. A. (2013). *ACS Chem. Biol.* **8**, 1981–1987.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.